

Détermination des modes de conduite des véhicules électriques dans des conditions réelles par des méthodes de classification

Mohamed BEN-MARZOUK^{1,2,*}, Guy CLERC¹, Serge PELISSIER², Ali SARI¹, Pascal VENET¹

¹Univ Lyon, Université Claude Bernard Lyon 1, Ecole Centrale de Lyon, INSA Lyon, CNRS, Ampère, F-69622, Villeurbanne, France

²Université de Lyon, IFSTTAR, AME, LTE, 69500 Bron, France

*mohamed.ben-marzouk@univ-lyon1.fr

RESUME – Pour étudier le vieillissement des batteries des véhicules électriques (VE), il faut déterminer précisément leurs conditions d'utilisation. En effet, à cause des spécificités des VE, ces usages sont très différents de ceux rencontrés dans les voitures thermiques

Dans cet article, nous présentons comment à partir des données d'usages réels, nous déterminons les différents modes de roulage. Nous présentons les différentes techniques utilisées pour analyser et classer les données et les différents modes d'usages obtenus.

Mots-clés— Véhicule électrique ; usage réel ; modes de roulage ; modes d'usage ; classification ; batterie.

1. INTRODUCTION

Grâce à la prise de conscience, de plus en plus importante chez les citoyens, sur les risques environnementaux et sur l'épuisement des ressources fossiles, et avec l'aide de la politique de plusieurs pays visant à baisser la pollution en ville, le véhicule électrique (VE) se trouve aujourd'hui en plein essor avec un nombre d'immatriculations qui ne cesse d'augmenter (hausse de 23% entre 2015 et 2016) [1]. Malgré cela, les VE restent néanmoins peu répandus principalement à cause des verrous technologiques liés aux batteries. En effet, celles-ci ont une autonomie faible, une durée de vie limitée et un prix actuellement élevé qui peut atteindre 40% de la valeur de la voiture.

Le problème de l'autonomie semble être en cours de résolution puisque les nouveaux modèles de VE affichent une autonomie plus importante que celle des modèles sortis il y a quelques années. De plus, le prix élevé des batteries peut être atténué par une production massive engendré par la démocratisation du VE. Cependant, le problème de la durée de vie reste toujours d'actualité. En effet, maîtriser la durée de vie des systèmes de stockage permet d'économiser sur le coût de remplacement en fin de vie et aussi sur le juste dimensionnement énergétique du système.

Il existe plusieurs travaux de recherche à travers le monde portant sur le vieillissement des batteries lithium-ion [2]–[4]. Cependant, les prédictions de durée de vie s'appuient souvent soit sur des modèles et des résultats issus d'essais de cyclage normalisés avec des profils de vieillissement accéléré constitués

par des charges/décharges, partielles ou totales, à courants constants, soit sur des profils simplifiés inspirés d'usages réels mais qui sont composés d'impulsions simples et ne représentent que partiellement la diversité des usages.

Pour étudier les problématiques liées à la durée de vie limitée des batteries en usage automobile, il est délicat de se fier aux simulations sur des cycles normalisés tels que le NEDC (New European Driving Cycle) ou le WLTC (Worldwide Harmonized Light Vehicles Test Cycles) car ils sont conçus pour les véhicules thermiques et donc leurs représentativités sont questionnables. Les VE et leurs fonctionnements, dépendent de nombreux facteurs et paramètres spécifiques, et pas seulement de la vitesse. Ils dépendent, en effet, du comportement de la batterie donc, entre autres, de la température et de son état de charge (SoC).

C'est pour cette raison que nous proposons de chercher, à partir des relevés réels des sollicitations des batteries et leurs conditions de fonctionnement, les différents modes d'usage des VE et leurs effets sur le vieillissement des batteries.

Dans cet article, nous présentons la méthodologie suivie pour chercher et identifier les différents modes d'usage des véhicules électriques. Tout d'abord, la base de données utilisée est présentée. Ensuite, les différentes étapes réalisées pour obtenir les différents modes d'usages sont détaillées.

2. DESCRIPTION DES DONNEES

Cette étude se base sur des données réelles, extraites de l'usage d'une dizaine de VE et issues du projet CROME (CROss-border Mobility for Electric vehicles) [5]. Plusieurs partenaires français et allemands ont travaillé, dans le cadre de ce projet, pour concevoir, réaliser et analyser un système de mobilité électrique entre la France et l'Allemagne. Ce projet a permis à l'IFSTTAR d'acquérir une très grande base de données relative à des VE en usage réel.

Ces VE, ayant la même architecture et appartenant à des volontaires particuliers et professionnels, ont été équipés avec des enregistreurs de données qui relèvent et transmettent, via une liaison GSM (Global System for Mobile communications), un grand nombre (quelques centaines) de variables liées au

fonctionnement (vitesse, appuis frein, enfoncement pédale de l'accélération, etc.), des données liées aux batteries (courant, tension, température des cellules, etc.) et d'autres liées au moteur (couple, température moteur, etc.).

Pour chacun des véhicules, un suivi de près de 2 ans a permis de générer des milliers de fichiers. Chaque fichier représente un usage du VE en roulage ou en charge. Dans ces fichiers, nous trouvons un suivi de l'évolution de toutes les variables liées au VE au cours du temps avec une fréquence d'échantillonnage de 10 Hz.

Chaque VE possède un pack batterie de 16 kWh ayant une tension nominale continue de 325,6 V. Ces packs batterie sont constitués de 88 cellules en série de type LMO (lithium manganèse spinel). Ces cellules ont une capacité de 50A h et une tension nominale de 3,7 V.

Comme mentionné précédemment, nous avons des centaines de variables à étudier. Dans la détermination des différents modes d'usage, certaines variables ont plus d'importance que d'autres. C'est pour cette raison que nous avons procédé à la sélection de paramètres que nous allons décrire dans la partie suivante.

3. SELECTION DES VARIABLES SIGNIFICATIVES

Pour sélectionner les variables les plus significatives sans perdre d'information, les variables qui n'ont aucun lien avec les modes de fonctionnement des batteries ont été éliminées (comme la température de l'eau de refroidissement du moteur par exemple).

Ainsi, le nombre de variables à suivre a été réduit de plus de 500 à une centaine. Néanmoins, comme ce nombre de variables restait toujours important, nous avons utilisé les méthodes indiquées ci-dessous pour réduire leur nombre en limitant la perte d'informations.

3.1. Étude d'intercorrélation

L'objectif de cette manipulation est d'enlever les redondances et de garder qu'une seule variable sur un ensemble inter-corrélé. Nous calculons, dans cette partie, les coefficients de corrélation entre les variables.

Pour étudier les relations entre les variables, nous utilisons le coefficient de Pearson « r » (corrélation linéaire) et le coefficient de Spearman « ρ » (corrélation des rangs).

Selon S. Tufféry [6], même si les variables sont continues, « on a toujours intérêt à comparer les deux coefficients de Pearson et Spearman, le plus fiable étant le second », notamment pour détecter des liaisons non linéaires.

Dans les paragraphes suivants nous allons présenter ces deux coefficients de corrélations et les résultats obtenus.

3.1.1. Coefficient de corrélation linéaire " r " de Pearson

Ce coefficient permet d'étudier l'existence d'une relation linéaire entre deux variables X et Y . Il s'obtient en divisant la covariance entre X et Y (σ_{XY}) par le produit de leurs écart-type, respectivement σ_X et σ_Y .

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}} \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}}} \quad (1)$$

Avec n le nombre d'observations, \bar{X} la moyenne de X et \bar{Y} la moyenne de Y .

3.1.2. Coefficient de corrélation de Spearman " ρ "

Nous avons aussi utilisé le coefficient de corrélation de Spearman, qui permet d'identifier une liaison monotone, linéaire ou non, entre les variables considérées. Cette technique étudie la relation entre les rangs de deux variables. Autrement dit nous n'utilisons pas les valeurs des observations dans les calculs mais leur rang. Si nous considérons une variable $X = (10, 13, 12, 22, 25, 8)$, X contient 6 observations, pour retrouver le vecteur rang_X , nous commençons par trier la variable X suivant un ordre croissant. Ensuite, nous remplaçons les valeurs que prend la variable X par leur numéro d'ordre. Ce qui donne $\text{rang}_X = (2, 4, 3, 5, 6, 1)$.

Pour calculer le coefficient de Spearman pour X et Y , nous pouvons utiliser l'équation (1) en remplaçant X par rang_X et Y par rang_Y . Ce coefficient peut aussi être donné par l'équation suivante.

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} \quad (2)$$

Avec n le nombre d'observations et d_i la différence entre rang_{X_i} et rang_{Y_i}

3.1.3. Condition de corrélation

Plusieurs variables sont fortement corrélées entre elles. Nous avons considéré dans la suite que deux variables X et Y sont fortement corrélées si elles suivent la condition suivante :

$$(r_{(X,Y)} > 0.75 \text{ ET } \rho_{(X,Y)} > 0.8) \quad (3)$$

Cette méthode permet d'avoir 18 groupes de variables inter-corrélées. Pour le reste de cette étude, nous choisissons une seule variable équivalente parmi un ensemble corrélé. Pour choisir correctement les variables à conserver, nous avons calculé et comparé les quantités d'informations contenues dans les variables en utilisant l'entropie de Shannon.

3.2. Sélection des variables

3.2.1. Entropie de Shannon

L'entropie de Shannon est une fonction mathématique qui peut mesurer la quantité d'information dans un signal. En effet, plus l'entropie d'un signal est grande, plus elle contient des informations non redondantes [7].

Considérons X un signal de m individus, contenant n symboles différents avec ($n < m$). Exemple : $X = (a; b; a; c; a; b; d)$, comme m est le nombre d'individus alors $m=7$ et dans le vecteur X il existe 4 symboles différents ("a", "b", "c" et "d") alors $n=4$. Pour mesurer la quantité d'information il faut tout d'abord calculer la probabilité d'apparence de chacun des symboles i avec ($i \in n$), nous l'appelons P_i . Pour notre exemple $P = (\frac{3}{7}; \frac{2}{7}; \frac{1}{7}; \frac{1}{7})$. Quant au calcul de l'entropie, nous appliquons l'équation (4). Pour notre exemple, nous obtenons $H(X) = 1.8482$.

Avec P_i la probabilité d'apparence de chacun des symboles i avec ($i \in n$), l'entropie de chacune des variables est donnée par l'équation suivante :

$$H(X) = - \sum_{i=1}^n P_i \log_2(P_i) \quad (4)$$

Avant l'application de cette fonction, les données sont normalisées. Cette normalisation pour chacune des variables est calculée suivant l'équation (5). Nous appelons X_{norm} le vecteur normalisé de X . Pour chaque variable, la valeur maximale est égale à 1 et la valeur minimale 0. Nous ne conservons que 4 décimales pour X_{norm} .

$$X_{norm} = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (5)$$

Lorsque nous avons un ensemble de variables corrélées, nous ne gardons que celui ayant l'entropie la plus élevée.

3.3. Résultats de l'étude de la corrélation

Cette méthodologie nous a permis de réduire le nombre de variables à étudier à 18 variables d'entrée non corrélées ou faiblement corrélées. La liste des variables sélectionnées et une brève description de chacune d'elles sont présentés dans le Tableau 1.

Tableau 1. Liste des variables sélectionnées

Nom de la variable	Description	Unité
SoC initial	État de charge en début de roulage en %	
Courant charge maxi	Courant maximal de la batterie en phase de charge	A
Puissance charge moyenne	Puissance moyenne de la batterie en phase de charge	W
Décélération moyenne	Décélération moyenne du véhicule	m.s ⁻²
Q dech par km	Quantité de charge de la batterie dissipée par km	Ah/km
Q charge	Quantité de charge de la batterie récupérée	Ah
Distance	Distance du véhicule	km
Q decharge	Quantité de charge de la batterie dissipée	Ah
Vitesse moyenne	Vitesse moyenne du véhicule pendant le roulage	km/h
Vitesse maxi	Vitesse maximale du véhicule	km/h
courant de decharge maxi	Courant maximal de la batterie en phase de décharge	A
Courant efficace	Courant efficace de la batterie	A
Puissance de decharge moy	Puissance moyenne de la batterie en phase de décharge	W
RPA	Relative Positive Acceleration (RPA) du véhicule	m.s ⁻²
Accélération moyenne	Accélération moyenne du véhicule	m.s ⁻²
PKE	Positive acceleration Kinetic Energy (PKE) du véhicule	m.s ⁻²
Q auxiliaires	Quantité de charge de la batterie consommée par les auxiliaires	Ah
Température ambiante	Température ambiante	°C

PKE et RPA sont des paramètres liés à l'éco-conduite qui sont liés à l'agressivité de la conduite [8], [9]

Toutes ces variables peuvent permettre de résumer comment et dans quelles conditions sont utilisées les batteries. Néanmoins, ce nombre de variables reste encore trop important

pour réaliser une classification efficiente capable de déterminer les différents modes d'usage.

4. REDUCTION DES DIMENSIONS

Le nombre de variables étant toujours élevé, nous avons appliqué la méthode de l'Analyse en Composantes Principales (ACP) afin de le réduire en projetant les variables sur les axes de l'ACP tout en déformant le moins possible la réalité [10].

4.1. Définitions

L'ACP est une méthode fondamentale dans les statistiques descriptives multidimensionnelles. Elle permet le traitement simultané de n'importe quel nombre de variables.

L'ACP projette des données sur des axes orthogonaux, ce qui signifie qu'il s'agit d'une transformation de variables pouvant être corrélées, à de nouvelles variables non corrélées. Ce processus réduit non seulement le nombre de variables mais rend également l'information moins redondante.

Le choix de dimension, autrement dit le nombre de composantes à considérer est une étape très importante. Elle dépend principalement de la qualité de la projection des observations et des variables sur les axes de l'ACP.

4.2. Définitions

Pour bien choisir le nombre de composantes à retenir, il y a plusieurs pratiques adaptées. Nous citons ici le critère de coude (rupture de pente) et la loi de Kaiser [11] qui propose de ne retenir que les axes qui ont une inertie supérieure ou égale à l'inertie moyenne[12]. L'inertie totale d'un nuage de points est la somme des carrés des distances des points par rapport au centre de gravité. L'inertie moyenne est l'inertie totale divisée par le nombre de variables. Lorsque nous projetons les données sur les axes de l'ACP, nous pouvons calculer l'inertie expliquée par chacune des composantes principales. Ce calcul va nous permettre d'appliquer ces deux critères pour le choix du nombre de composantes principales.

Le graphe suivant (Fig. 1) montre l'inertie expliquée par chaque axe de l'ACP par rapport à l'inertie totale.

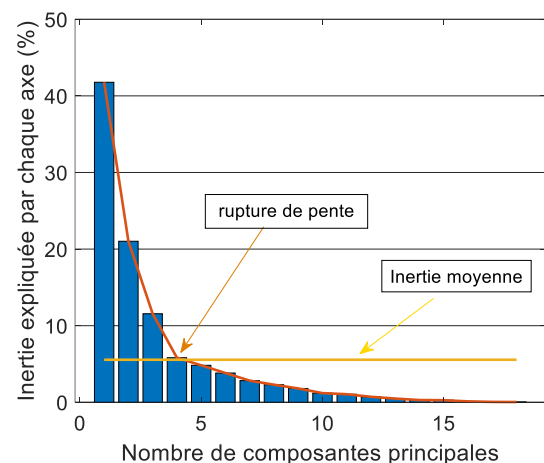


Fig. 1. Inertie pour chaque composante principale

Dans notre cas, en suivant les 2 critères cités précédemment, nous pouvons nous arrêter à 4 axes et considérer 80% de l'inertie totale. Il faut tout de même vérifier si avec seulement 4 axes la qualité de la projection des variables est acceptable. La qualité de la projection des variables est calculée de la façon suivante [6], [13]:

$$QV_n = \sum_{i=1}^n \text{corr}^2(V, Ax_i) \quad (6)$$

avec QV_n la qualité de projection de la variable "V" sur "n" axes d'ACP. corr^2 est le carré de la corrélation de Pearson. V représente le vecteur des données de la variable "V". Ax_i est le vecteur des données du $i^{\text{ème}}$ axe de l'ACP.

Le tableau suivant illustre la qualité de la projection des variables par rapport au nombre de composantes principales. La couleur rouge signifie une mauvaise qualité de projection et la couleur verte une bonne qualité de projection.

Tableau 2. Qualité de projection des variables en fonction du nombre de composantes principales

	Nombre d'axes (composantes principales)						
	1	2	3	4	5	6	7
SoC initial	1%	1%	1%	95%	95%	99%	99%
Courant charge maxi	43%	51%	53%	53%	68%	75%	77%
Puissance charge moyenne	47%	66%	69%	70%	70%	75%	75%
Deceleration moyenne	0%	84%	84%	84%	87%	94%	94%
Q dech par km	1%	3%	74%	74%	78%	80%	89%
Q charge	52%	52%	60%	60%	76%	84%	90%
Distance	61%	86%	86%	88%	88%	94%	94%
Q decharge	68%	87%	88%	90%	91%	97%	97%
Vitesse moyenne	74%	79%	81%	81%	96%	96%	97%
Vitesse maxi	85%	86%	87%	88%	88%	89%	89%
courant de decharge maxi	66%	68%	73%	75%	81%	83%	83%
Courant efficace	81%	92%	94%	94%	95%	96%	97%
Puissance de decharge moy	80%	83%	85%	85%	94%	94%	97%
RPA	41%	91%	92%	92%	96%	96%	96%
Acceleration moyenne	0%	88%	88%	89%	89%	94%	94%
PKE	9%	79%	82%	83%	92%	92%	92%
Q auxiliaires	17%	25%	75%	76%	77%	81%	83%
Temperature ambiante	1%	1%	57%	59%	67%	67%	99%

Les deux critères cités précédemment recommandent un choix de 4 axes d'ACP. Mais selon le Tableau 2, il est plus judicieux de choisir 7 composantes principales parmi 18 (nombre de variables) pour garantir une qualité de projection supérieure à 75% pour chacune des variables.

Cette étape nous a donc permis de réduire la dimension et d'éliminer les redondances ce qui permet de réaliser une meilleure identification des modes de fonctionnement.

5. IDENTIFICATION DES MODES DE CYCLES DE CONDUITE

Une fois le nombre de variables réduit au maximum, nous pouvons étudier les cycles de conduite (roulages) définis entre 2 arrêts « moteur » du VE. Comme nous avons plusieurs milliers de roulages enregistrés, nous nous sommes intéressés à identifier les similitudes et les non-similitudes entre ces roulages. Nous cherchons ainsi à obtenir plusieurs classes de cycles de conduite en faisant appel aux techniques de classification.

5.1. Méthodes de classification

Les méthodes de classification peuvent être divisées en deux familles principales :

- la classification supervisée où des éléments sont déjà classés et où l'ajout de nouveaux éléments est à effectuer,
- la classification non supervisée lorsque nous n'avons pas de connaissance préalable des classes.

Dans cette étude, puisque nous ne disposons pas de connaissances préalables sur les classes, nous utilisons une méthode de classification non supervisée pour identifier les modes d'usage des VE. La classification non supervisée est une technique de traitement des données qui vise à classer un

ensemble de données en minimisant la distance intra-classe et en maximisant autant que possible la distance entre les classes. Il existe de nombreuses techniques de mesure de distance[14] comme la distance de Manhattan et la distance euclidienne. Dans cette partie, nous utilisons le carré de la distance euclidienne.

Il existe plusieurs méthodes de classification non supervisée, les plus connues étant la Classification Ascendante Hiérarchique (CAH) et les « k-means » [6].

5.1.1. Classification Ascendante Hiérarchique (CAH)

La CAH regroupe itérativement les individus en agrégeant 2 par 2, les éléments les plus proches qui permettent de construire progressivement un dendrogramme (diagramme en arbre) qui regroupe à la fin tous les individus d'une même classe. Sur la Fig. 2, nous présentons un exemple qui illustre l'algorithme de CAH. L'algorithme commence par regrouper les observations Ob1 et Ob2 car ce sont les éléments les plus proches. Ensuite, Ob3 et Ob4 sont regroupés. La fin du calcul est atteinte lorsque toutes les observations sont réunies dans un même groupe.

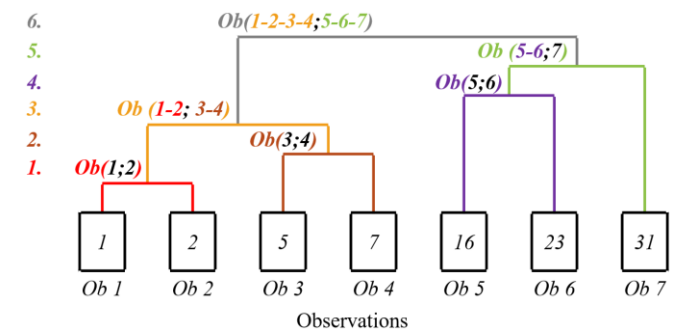


Fig. 2. Illustration de l'algorithme de CAH

5.1.2. K-means

La méthode des « k-means » consiste à regrouper les observations en k groupes de sorte que les distances intra-classes soient minimales et les distances entre classes maximales. L'algorithme « k-means » fonctionne comme suit :

Après avoir initialisé k points (aléatoirement ou non) et en considérant leurs centroïdes (centres de gravité), l'algorithme répartit les points (observations) dans les k classes ainsi formées en fonction de leur proximité avec le centroïde. Ensuite, l'algorithme recalcule les centres de gravité des classes et les considère comme de nouveaux centroïdes. L'algorithme répète ces 2 étapes jusqu'à ce qu'il n'y ait plus de changement. La figure 4 illustre le fonctionnement de l'algorithme « k-means ».

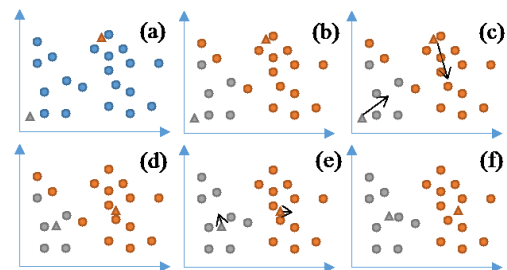


Figure 3: Illustration de l'algorithme des « k-means »

La figure 4 (a) montre l'initialisation des centroïdes (marqueurs triangulaires), les points sont en bleu pour indiquer qu'ils n'appartiennent encore à aucune classe. La figure 4 (b) montre comment les points sont classés parmi 2 classes (grise et orange). Les figures 4 (c) et (d) présentent l'étape du calcul des nouveaux centroïdes. La figure 4 (e) représente une autre itération des étapes précédentes. Enfin, la figure 4 (f) montre le résultat final de cette classification.

5.2. Classification des roulages

Pour classer ces roulages, nous avons choisi la méthode des « k-means » car elle est plus rapide. A l'opposé de la CAH, elle effectue une mise à jour des centres des classes après chaque nouvelle affectation pour suivre l'évolution de son contenu. Dans la CAH, si deux individus sont placés dans des classes différentes, ils ne sont jamais comparés à nouveau.

Les résultats obtenus ont révélé que pour classer correctement les roulages avec un minimum de classes, nous devons en garder 5. Ainsi avec 5 classes seulement nous pouvons expliquer près de 60% de la variance (pour expliquer 100% de la variance il faudrait que le nombre de classes soit égal au nombre de roulages sachant que notre base de données comporte près de 8000 roulages).

La proportion de la variance expliquée R^2 est un indicateur de la qualité de la classification. Elle peut s'exprimer grâce à l'équation suivante :

$$R^2 = \frac{I_B}{I} = \frac{I_B}{I_B + I_W} \quad (3)$$

avec I l'inertie totale, I_B l'inertie interclasses et I_W l'inertie intra-classes. Ces deux inerties peuvent s'exprimer par :

$$I_B = \sum_i^k distance(C_i, C) \quad (4)$$

$$I_W = \frac{1}{n} \sum_i^k \sum_{l_{ji}}^{m_i} distance(l_{ji}, C_i) \quad (5)$$

k est le nombre de classes, C_i est le centre de la classe i , C représente le centre de gravité de tous les roulages, n est le nombre de roulages, m_i est le nombre de roulages appartenant à la classe i et l_{ji} est le roulage numéro j appartenant à la classe i .

5.3. Résultats obtenus

Après le classement des cycles de conduite en 5 catégories, nous avons considéré qu'un mode de fonctionnement est la moyenne des roulages appartenant à la même classe.

La Fig.4 montre la différence entre les modes de fonctionnement selon les 18 variables choisies précédemment.

Selon les variables « distance », « vitesse moyenne » et « vitesse maximale », nous pouvons constater que les modes 1 et 2 représentent les cycles de conduite urbains et les modes 4 et 5 les cycles de conduite qui incluent des parties d'autoroute. De plus, selon les variables « Accélération moyenne », « décélération moyenne », « RPA » (Relative Positive Acceleration - indicateur fondé sur l'accélération et la vitesse) et « PKE » (Positive Kinetic Energy - indicateur de variation de l'énergie cinétique), nous pouvons remarquer que le mode 4 décrit un comportement plus agressif comparé au mode 5. Il en est de même pour le mode 2 par rapport au mode 1.

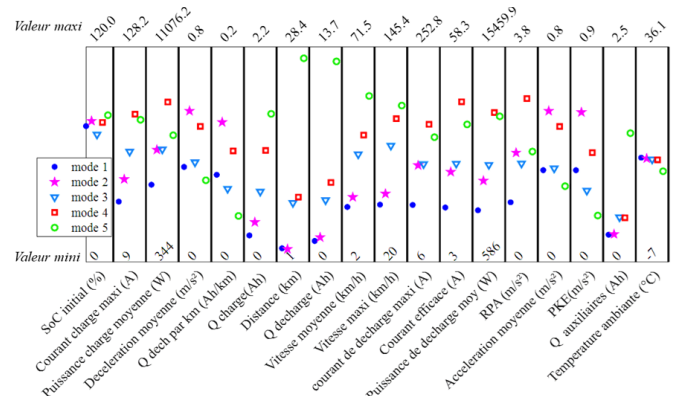


Fig. 4. Caractéristique de chaque mode d'usage

Cette méthodologie permet d'obtenir plusieurs classes de cycles de conduite qui dépendent de différentes variables. Nous allons par la suite étudier l'influence de ces différents modes de roulage sur le vieillissement des batteries.

6. CONCLUSION

Le présent travail montre comment à partir d'enregistrements réels sur des véhicules électriques et de sollicitations des batteries, nous déterminons différents modes de fonctionnement des batteries. En utilisant plusieurs outils de sélection basés sur les coefficients de corrélation de Pearson et de Spearman, sur l'entropie de Shannon, nous avons réduit le nombre de paramètres à 18. Pour une meilleure qualité de classification, nous avons utilisé la technique ACP qui a permis de réduire la dimension et d'éliminer les redondances. Une analyse de type « k-means » nous a permis de trouver 5 modes d'usages différents. Chaque mode est défini par les valeurs des 18 paramètres précédemment identifiés. Les outils développés dans ce travail permettront d'étudier les liens potentiels entre le taux de vieillissement des batteries et les modes d'utilisation identifiés.

7. REMERCIEMENTS

Nous remercions la région Auvergne-Rhône-Alpes pour son soutien financier pour ce travail.

8. REFERENCES

- [1] Avere-France, « véhicules électriques immatriculés en 2016 », *Avere-France*. [En ligne]. Disponible sur: http://www.aver-france.org/Site/Article/?article_id=6826.
- [2] S. B. Peterson, J. Apt, et J. F. Whitacre, « Lithium-ion battery cell degradation resulting from realistic vehicle and vehicle-to-grid utilization », *J. Power Sources*, vol. 195, n° 8, p. 2385-2392, avr. 2010.
- [3] E. Wood, M. Alexander, et T. H. Bradley, « Investigation of battery end-of-life conditions for plug-in hybrid electric vehicles », *J. Power Sources*, vol. 196, n° 11, p. 5147-5154, 2011.
- [4] E. Redondo-Iglesias, P. Venet, et S. Pelissier, « Eyring acceleration model for predicting calendar ageing of lithium-ion batteries », *J. Energy Storage*, vol. 13, p. 176-183, oct. 2017.
- [5] P. Kreczanik, B. Jeanneret, et S. Pelissier, « Construction of Database on Real World Uses of Electric Vehicles-A French Case », in *Vehicle Power and Propulsion Conference (VPPC), 2014 IEEE*, 2014, p. 1-5.
- [6] S. Tufféry, *Data mining et statistique décisionnelle - 4ème édition*, 4e édition. Paris: Editions technip, 2012.
- [7] R. M. Gray, *Entropy and information theory*, 2nd ed. New York: Springer, 2011.
- [8] K. S. Nesamani et K. P. Subramanian, « Development of a driving cycle for intra-city buses in Chennai, India », *Atmos. Environ.*, vol. 45, n° 31, p. 5469-5476, oct. 2011.
- [9] G. Castignani, T. Derrmann, R. Frank, et T. Engel, « Validation study of risky event classification using driving pattern factors », in *2015 IEEE Symposium on Communications and Vehicular Technology in the Benelux (SCVT)*, 2015, p. 1-6.
- [10] G. Saporta, *Probabilités, analyse des données et statistique*. Paris: Technip, 2011.

- [11] D. D. Suhr, « Principal component analysis vs. exploratory factor analysis », *SUGI 30 Proc.*, vol. 203, p. 230, 2005.
- [12] A. S. Beavers, J. W. Lounsbury, J. K. Richards, S. W. Huck, G. J. Skolits, et S. L. Esquivel, « Practical considerations for using exploratory factor analysis in educational research. », *Pract. Assess. Res. Eval.*, vol. 18, 2013.
- [13] R. R. Herrera et D. S.-L. Gac, *Initiation à l'analyse factorielle des données : Fondements des mathématiques et interprétations, cours et exercices*. Paris: Ellipses Marketing, 2002.
- [14] M. Bora, D. Jyoti, D. Gupta, et A. Kumar, « Effect of different distance measures on the performance of K-means algorithm: an experimental study in Matlab », *ArXiv Prepr. ArXiv14057471*, 2014.